



10/506454 PCT

## RAW SEQUENCE LISTING

DATE: 09/09/2004

PATENT APPLICATION: US/10/506,454

TIME: 16:09:18

Input Set : D:\FID001 US seq list.txt

Output Set: N:\CRF4\09092004\J506454.raw

3 <110> APPLICANT: Slesarev, Alexi I  
 4 Mezhevaya, Katja V  
 5 Polushin, Nikolai N  
 6 Shcherbinina, Olga V  
 7 Shakhova, Vera V  
 8 Malykh, Andrei G  
 9 Kozyavkin, Sergei A  
 11 <120> TITLE OF INVENTION: The Complete Genome and Protein Sequences of the  
 Hyperthermophile  
 12 Methanopyrus Kandleri AV19 and Monophyly of Archaeal Methanogens  
 13 and Methods of Use Thereof  
 15 <130> FILE REFERENCE: FID001  
 C--> 17 <140> CURRENT APPLICATION NUMBER: US/10/506,454  
 C--> 17 <141> CURRENT FILING DATE: 2004-08-31  
 17 <150> PRIOR APPLICATION NUMBER: PCT/US03/06664  
 18 <151> PRIOR FILING DATE: 2003-03-04  
 20 <150> PRIOR APPLICATION NUMBER: 60/361,742  
 21 <151> PRIOR FILING DATE: 2002-03-04  
 23 <160> NUMBER OF SEQ ID NOS: 1722  
 25 <170> SOFTWARE: PatentIn version 3.2  
 27 <210> SEQ ID NO: 1  
 28 <211> LENGTH: 352  
 29 <212> TYPE: PRT  
 30 <213> ORGANISM: Methanopyrus kandleri  
 32 <400> SEQUENCE: 1  
 34 Met Ser Leu Ile Glu Ile Asp Gly Ser Tyr Gly Glu Gly Gly Gln  
 35 1 5 10 15  
 38 Ile Leu Arg Thr Ala Val Gly Met Ser Ala Leu Thr Gly Glu Pro Val  
 39 20 25 30  
 42 Arg Ile Tyr Asn Ile Arg Ala Asn Arg Pro Arg Pro Gly Leu Ser His  
 43 35 40 45  
 46 Gln His Leu His Ala Val Lys Ala Val Ala Glu Ile Cys Asp Ala Glu  
 47 50 55 60  
 50 Cys Glu Gly Leu Glu Ile Gly Ser Thr Glu Ile Val Phe Glu Pro Gly  
 51 65 70 75 80  
 54 Lys Val Lys Gly Gly Glu Tyr Glu Val Asp Ile Gly Thr Ala Gly Ser  
 55 85 90 95  
 58 Val Thr Leu Leu Leu Gln Ala Val Lys Leu Ala Ala Ile Ala Ala Asp  
 59 100 105 110  
 62 Gly Pro Val Glu Met Glu Val Arg Gly Gly Thr Asp Val Lys Trp Ser  
 63 115 120 125  
 66 Pro Pro Val Asp Tyr Glu Ile Asn Val Asn Ala His Tyr Leu Asp Arg  
 67 130 135 140  
 70 Leu Gly Tyr Arg Tyr Glu Leu Glu Val Leu Arg Arg Gly His Tyr Pro



## RAW SEQUENCE LISTING

DATE: 09/09/2004

PATENT APPLICATION: US/10/506,454

TIME: 16:09:18

Input Set : D:\FID001 US seq list.txt

Output Set: N:\CRF4\09092004\J506454.raw

```

71 145          150          155          160
74 Arg Gly Gly Gly Ile Val Arg Ala Arg Met Glu Pro Pro Lys Arg Leu
75          165          170          175
78 Lys Pro Leu Glu Ala Val Lys Phe Gly Glu Leu Glu Ser Val Arg Gly
79          180          185          190
82 Ile Ser His Cys Val Arg Leu Pro Pro His Val Ala Glu Arg Gln Ala
83          195          200          205
86 Lys Ala Ala Ser Glu Ile Ile Glu Arg Glu Leu Gly Ile Arg Pro Glu
87          210          215          220
90 Ile Glu Ile Glu Thr Tyr Pro Lys Gly Arg Asp Pro His Leu Gly Pro
91 225          230          235          240
94 Gly Ser Gly Ile Val Leu Trp Ala Glu Asp Asp Gln Gly Asn Arg Ile
95          245          250          255
98 Gly Ala Asp Ala Leu Gly Glu Lys Gly Lys Pro Ala Glu Val Val Gly
99          260          265          270
102 Arg Glu Ala Ala Glu Gln Leu Val Gln Arg Leu Arg Thr Gly Met Ala
103          275          280          285
106 Leu Asp Glu His Met Gly Asp Gln Ile Leu Pro Phe Leu Ala Ile Ala
107          290          295          300
110 Asp Gly Glu Ser Val Phe Gly Val Thr Gly Val Asp Pro His Leu Pro
111 305          310          315          320
114 Thr Asn Ala Trp Val Val Glu Lys Phe Leu Pro Val Ser Val Glu Ile
115          325          330          335
118 Arg Gly Lys Glu Gly Glu Pro Ala Thr Val Glu Val Arg Pro Glu Gly
119          340          345          350
122 <210> SEQ ID NO: 2
123 <211> LENGTH: 171
124 <212> TYPE: PRT
125 <213> ORGANISM: Methanopyrus kandleri
127 <400> SEQUENCE: 2
129 Val Gly Val Ile Glu Asp Met Met Lys Val Gly Met Arg Ser Ala Lys
130 1          5          10          15
133 Ala Gly Leu Glu Ala Thr Glu Glu Leu Ile Lys Leu Phe Arg Glu Asp
134          20          25          30
137 Gly Arg Leu Val Gly Ser Ile Leu Lys Glu Met Glu Pro Glu Glu Ile
138          35          40          45
141 Thr Glu Leu Leu Glu Gly Ala Ser Ser Gln Leu Ile Arg Met Ile Arg
142          50          55          60
145 Ser Leu His Thr Pro Ala Val Asp Val Phe Glu Arg Ser Gly Glu Phe
146 65          70          75          80
149 Val Ile Val Ala Glu Val Pro Gly Ala Arg Pro Glu Asp Val Gln Val
150          85          90          95
153 Arg Ala Gly Glu Arg Phe Val Glu Ile Thr Ala Asn Ile Pro Lys Met
154          100          105          110
157 Arg Glu Gly Glu Ala Lys Thr Arg Glu Arg Val Thr Gly Glu Val Arg
158          115          120          125
161 Arg Arg Ile Asp Leu Pro Glu Lys Ile Asn Pro Ser Ala Val Ser Ala
162          130          135          140
165 Lys Cys Gly Arg Gly Leu Leu Ile Val Arg Ala Pro Lys Ala Glu Ala

```

## RAW SEQUENCE LISTING

DATE: 09/09/2004

PATENT APPLICATION: US/10/506,454

TIME: 16:09:18

Input Set : D:\FID001 US seq list.txt

Output Set: N:\CRF4\09092004\J506454.raw

```

166 145          150          155          160
169 Ser Glu Ile Glu Val Lys Pro Met Glu Glu Glu
170          165          170
173 <210> SEQ ID NO: 3
174 <211> LENGTH: 352
175 <212> TYPE: PRT
176 <213> ORGANISM: Methanopyrus kandleri
178 <400> SEQUENCE: 3
180 Val Ser Gly Asn Pro Phe Arg Lys Met Pro Glu Val Pro Asp Pro Glu
181 1          5          10          15
184 Glu Leu Ile Asp Val Ala Phe Arg Arg Ala Glu Arg Ala Ala Glu Gly
185          20          25          30
188 Thr Arg Lys Ser Phe Tyr Gly Thr Arg Thr Pro Pro Glu Val Arg Ala
189          35          40          45
192 Arg Ser Ile Glu Ile Ala Arg Val Asn Thr Ala Cys Gln Leu Val Gln
193 50          55          60
196 Asp Arg Leu Trp Glu Ile Val Arg Lys Thr Pro Asn Leu Asp Glu Leu
197 65          70          75          80
200 His Pro Phe Tyr Arg Glu Leu Ala Asp Ala Leu Ala Gly Ile Asp Arg
201          85          90          95
204 Leu Lys Ser Ser Leu Ala Asp Val His Thr Val Ala Lys Ile Ala Arg
205          100          105          110
208 Leu Ile Arg Glu Glu Tyr Thr Arg Lys Ile Lys Arg Ala Arg Asp Pro
209          115          120          125
212 Arg Thr Ala Ala Glu Leu Arg Arg Gln Ala Phe Gly Arg Leu Ala Ser
213          130          135          140
216 Thr Ile Arg Arg Lys Val Gly Asp Ala Leu Arg Phe Leu Arg Lys Val
217 145          150          155          160
220 Gln Pro Lys Leu Val Asp Leu Pro Ala Ile Asp Thr Glu Met Phe Thr
221          165          170          175
224 Val Thr Leu Ala Gly Phe Pro Asn Val Gly Lys Thr Thr Leu Met Thr
225          180          185          190
228 Val Leu Thr Gly Ser Arg Pro Glu Ile Ala Pro Tyr Pro Phe Thr Thr
229          195          200          205
232 Lys Gly Ile Gln Val Gly Tyr Met Glu Arg Pro Tyr Pro Val Gln Met
233          210          215          220
236 Leu Asp Thr Pro Gly Leu Leu Glu Arg Pro Glu Glu Glu Arg Asn Pro
237 225          230          235          240
240 Val Glu Arg Gln Ala Ile Ala Ala Leu Lys His Val Thr Asp Ala Val
241          245          250          255
244 Leu Phe Leu Ile Asp Pro Thr Gly Thr Cys Gly Tyr Pro Val Glu Glu
245          260          265          270
248 Gln Leu Glu Leu Leu Asp Arg Val Arg Lys Glu Phe Asp Val Pro Val
249          275          280          285
252 Tyr Val Val Leu Thr Lys Ala Asp Leu Arg Asp Leu Trp Glu Glu Pro
253          290          295          300
256 Asp Leu Glu Gly Glu Pro Val Tyr Lys Val Ser Ala Thr Glu Arg Thr
257 305          310          315          320
260 Gly Leu Lys Glu Leu Arg Glu Leu Leu Asn Asp Leu Ala Arg Gly His

```

## RAW SEQUENCE LISTING

DATE: 09/09/2004

PATENT APPLICATION: US/10/506,454

TIME: 16:09:18

Input Set : D:\FID001 US seq list.txt

Output Set: N:\CRF4\09092004\J506454.raw

```

261          325          330          335
264 Tyr Ser Gly Arg Asp Arg Gly His Asp Glu Gly Arg Asp Glu Glu Arg
265          340          345          350
268 <210> SEQ ID NO: 4
269 <211> LENGTH: 105
270 <212> TYPE: PRT
271 <213> ORGANISM: Methanopyrus kandleri
273 <400> SEQUENCE: 4
275 Met Glu Tyr Ile Tyr Ala Ala Leu Leu Leu His Ala Ala Gly Gln Glu
276 1          5          10          15
279 Ile Asn Glu Asp Asn Leu Arg Lys Val Leu Glu Ala Ala Gly Val Asp
280          20          25          30
283 Val Asp Asp Ala Arg Leu Lys Ala Thr Val Ala Ala Leu Glu Glu Val
284          35          40          45
287 Asp Ile Asp Glu Ala Ile Glu Glu Ala Ala Val Pro Ala Ala Ala Pro
288          50          55          60
291 Ala Ala Ala Ala Pro Ala Glu Glu Glu Glu Glu Glu Ala Glu Ala
292 65          70          75          80
295 Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Ala Glu Glu Glu Ala
296          85          90          95
299 Ala Ala Gly Leu Gly Ala Leu Phe Gly
300          100          105
303 <210> SEQ ID NO: 5
304 <211> LENGTH: 510
305 <212> TYPE: PRT
306 <213> ORGANISM: Methanopyrus kandleri
308 <400> SEQUENCE: 5
310 Leu Val Pro Trp Val Glu Lys Tyr Arg Pro Arg Ser Leu Lys Glu Leu
311 1          5          10          15
314 Val Asn Gln Asp Glu Ala Lys Lys Glu Leu Ala Ala Trp Ala Asn Glu
315          20          25          30
318 Trp Ala Arg Gly Ser Ile Pro Glu Pro Arg Ala Val Leu Leu His Gly
319          35          40          45
322 Pro Pro Gly Thr Gly Lys Thr Ser Ala Ala Tyr Ala Leu Ala His Asp
323          50          55          60
326 Phe Gly Trp Asp Val Ile Glu Leu Asn Ala Ser Asp Lys Arg Thr Arg
327 65          70          75          80
330 Asn Val Ile Glu Lys Ile Val Gly Gly Ala Ser Thr Ser Arg Ser Leu
331          85          90          95
334 Leu Arg Met Thr Arg Glu Ala Gly Gly Asp Tyr Glu His Val Glu Gly
335          100          105          110
338 His Ser Asp Arg Val Leu Val Leu Val Asp Glu Val Asp Gly Ile Asp
339          115          120          125
342 Pro Arg Glu Asp Arg Gly Gly Val Thr Ala Leu Thr Arg Ala Val Arg
343          130          135          140
346 Gln Ala Arg Asn Pro Met Val Leu Val Ala Asn Asp Pro Trp Val Leu
347 145          150          155          160
350 Pro Lys Ser Leu Arg Asp Ala Val Arg Met Ile Glu Phe Arg Arg Leu
351          165          170          175

```

## RAW SEQUENCE LISTING

DATE: 09/09/2004

PATENT APPLICATION: US/10/506,454

TIME: 16:09:18

Input Set : D:\FID001 US seq list.txt

Output Set: N:\CRF4\09092004\J506454.raw

```

354 Arg Val Asn Asp Ile Val Glu Ala Leu Arg Arg Ile Cys Glu Arg Glu
355           180           185           190
358 Gly Ile Glu Tyr Glu Glu Val Ala Leu Arg Arg Ile Ala Lys Arg Ala
359           195           200           205
362 Arg Gly Asp Leu Arg Ala Ala Ile Asn Asp Leu Glu Ala Leu Ala Arg
363           210           215           220
366 Pro Thr Gly Arg Val Thr Ser Asp Asp Val Glu Ala Leu Gly Trp Arg
367 225           230           235           240
370 Asp Lys Glu Ile Thr Ile Phe Glu Ala Leu Gly Arg Ile Phe Asn Lys
371           245           250           255
374 Pro Pro Arg Gln Ala Arg Arg Ala Leu Trp Asn Leu Asp Glu Asp Pro
375           260           265           270
378 Asp Asp Val Ile Leu Trp Ile Ala Gln Asn Ile Pro Arg Ala Tyr Arg
379           275           280           285
382 Asp Pro Glu Glu Ile Ala Arg Ala Tyr Asp Tyr Leu Ser Lys Ala Asp
383           290           295           300
386 Val Phe Ser Ser Arg Ala Ile Glu Thr Gly Asp Trp Arg Phe Lys Tyr
387 305           310           315           320
390 Val Tyr Ala Thr Asp Leu Met Thr Ser Gly Val Ala Ala Ala Arg Lys
391           325           330           335
394 Gly Lys Pro Pro Gly Phe Val Arg Phe Gln Pro Pro Lys Ile Leu Arg
395           340           345           350
398 Lys Leu Gly Thr Thr Arg Lys Glu Arg Glu Val Arg Asn Ser Ile Ala
399           355           360           365
402 Lys Lys Ile Ala Glu Arg Met His Val Ser Thr Arg Arg Ala Lys Met
403           370           375           380
406 Asp Val Ile Ser Val Leu Glu Ile Ala Phe Arg Lys Val Ala Asp Asn
407 385           390           395           400
410 Pro Thr Asp Arg Gly Leu Glu Ile Leu Gly Gly Ile Ala Gly Tyr Leu
411           405           410           415
414 Glu Leu Ser Lys Arg Glu Ile Gly Phe Leu Cys Gly Asp Pro Gln Val
415           420           425           430
418 Ala Gln Arg Val Tyr Gln Arg Ala Leu Arg Val Arg Glu Lys Leu Arg
419           435           440           445
422 Lys Ile Arg Arg Glu Arg Val Lys Gly Ala Met Glu Ser Met Leu Glu
423           450           455           460
426 Arg Lys Arg Glu Glu Ser Glu Val Glu Glu Glu Ala Lys Glu Ile Glu
427 465           470           475           480
430 Glu Ala Val Glu Lys Ala Glu Glu Glu Glu Glu Arg Glu Glu Lys Lys
431           485           490           495
434 Lys Glu Gly Gly Gly Glu Gln Arg Thr Leu Asp Ala Phe Phe
435           500           505           510
438 <210> SEQ ID NO: 6
439 <211> LENGTH: 635
440 <212> TYPE: PRT
441 <213> ORGANISM: Methanopyrus kandleri
443 <400> SEQUENCE: 6
445 Met Ala Glu His Glu Leu Arg Val Leu Glu Ile Pro Trp Val Glu Lys
446 1           5           10           15

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/506,454

DATE: 09/09/2004  
TIME: 16:09:19

Input Set : D:\FID001 US seq list.txt  
Output Set: N:\CRF4\09092004\J506454.raw

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:249; Xaa Pos. 200  
Seq#:314; Xaa Pos. 200  
Seq#:315; Xaa Pos. 143,212  
Seq#:321; Xaa Pos. 132  
Seq#:627; Xaa Pos. 15,66  
Seq#:809; Xaa Pos. 22  
Seq#:931; Xaa Pos. 388  
Seq#:1342; Xaa Pos. 169,170,171,172  
Seq#:1361; Xaa Pos. 92,93,94,95,96,97,98,99  
Seq#:1367; Xaa Pos. 16  
Seq#:1410; Xaa Pos. 179  
Seq#:1525; Xaa Pos. 120  
Seq#:1547; Xaa Pos. 360,361,362

## VERIFICATION SUMMARY

DATE: 09/09/2004

PATENT APPLICATION: US/10/506,454

TIME: 16:09:19

Input Set : D:\FID001 US seq list.txt

Output Set: N:\CRF4\09092004\J506454.raw

L:17 M:270 C: Current Application Number differs, Replaced Current Application No  
L:17 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:21013 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:249 after pos.:192  
L:26282 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:314 after pos.:192  
L:26363 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:315 after pos.:128  
M:341 Repeated in SeqNo=315  
L:26691 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:321 after pos.:128  
L:50047 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:627 after pos.:0  
M:341 Repeated in SeqNo=627  
L:65939 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:809 after pos.:16  
L:76567 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:931 after pos.:384  
L:112286 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1342 after pos.:160  
L:113909 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1361 after pos.:80  
M:341 Repeated in SeqNo=1361  
L:114697 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1367 after pos.:0  
L:118548 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1410 after pos.:176  
L:126931 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1525 after pos.:112  
L:129383 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1547 after pos.:352